NBSIR 73-131

Human Factors Evaluation of a Voice Encoding System

V. J. Pezoldt
J. J. Persensky

Human Factors Laboratory
Social and Engineering Sciences Group
Technical Analysis Division
Institute for Applied Technology
National Bureau of Standards

March 1973

Final Report

Prepared for:
Postal Service Laboratory
Research Department
U. S. Postal Service
11711 Parklawn Drive
Rockville, Maryland 20852

NBSIR 73-131

HUMAN FACTORS EVALUATION OF A VOICE ENCODING SYSTEM

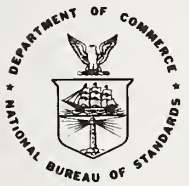
V. J. Pezoldt
J. J. Persensky

Human Factors Laboratory
Social and Engineering Sciences Group
Technical Analysis Division
Institute for Applied Technology
National Bureau of Standards
Washington, D. C. 20234

March 1973

Final Report

Prepared for:
Postal Service Laboratory
Research Department
U. S. Postal Service
11711 Parklawn Drive
Rockville, Maryland 20852



U. S. DEPARTMENT OF COMMERCE, Frederick B. Dent, Secretary

NATIONAL BUREAU OF STANDARDS, Richard W. Roberts, Director

TABLE OF CONTENTS

Executives' Summary

Human Factors Evaluation of a Voice Encoding System

Problem...

To evaluate a Voice Encoding System (VES)

- In a simulated parcel sorting condition (Phase I)
- To determine man-VES interface limitation (Phase II).

To study this problem...

Phase I...

4 naive subjects

- worked for 12 days each
- speaking the last 2 digits of the ZIP Code into the VES
- addresses were presented via LSM console at presentation rate of

  20 LPM and

  35 LPM
- under work/rest conditions of

  15/15 min.

  30/30 min.

  60/60 min.
- and in a fatigue test where they worked without rest until they reported fatigue.

Phase II...

3 highly motivated subjects

- worked in a 30/30 min. work/rest condition
- at 20 LPM
- to a criterion of

  95% accuracy or

  8 non-criterion trials
- reading

  1, 2, 3, and 4 digits of the ZIP Code, and
- at 15 LPM
- reading

  the last 2 digits of the ZIP Code
- with a cancel and correct capability available.

6 subjects

- 100 single digits from a list
- under 2 conditions of ambient noise

  LSM console on

  LSM console off
- this task was replicated 3 times in each condition.

Executives' Summary

Data collected were...

Phase I...

- Accuracy
  - System
  - Reading
- time to fatigue

Phase II...

- Accuracy
  - System
  - Reading
- Detailed command error analysis

Results...

Phase I...

- Subjects performed better
  - at slower speeds and
  - with longer work/rest cycles
- Average time to fatigue, without rest, was 2 hrs. 34 min.
- Naive subjects generally perform poorly
  - motivation and
  - training

  are necessary for this task.

Phase II...

- Motivated subjects' performance was 15% higher than Phase I subjects.

- Cancel capability resulted in 97% recognition accuracy.
- LSM noise did not affect recognition.

Conclusions and Recommendations...

- Voice encoding (VE) would be of limited value to high speed parcel sorting unless
  - used only during slack periods or
  - current facer and keyer both use VES.
- VE could be applied to other USPS tasks.
- Other VE systems should be tested.
- These systems should have
  - single command training
  - other than voice activation of cancel
  - require infrequent system training.
- Subjects should be
  - highly motivated
  - well trained.
- VES should be tested with subjects actually sorting parcels.
- Other USPS tasks should be analyzed for application of VE.

HUMAN FACTORS EVALUATION OF A VOICE ENCODING SYSTEM

1.0 BACKGROUND AND STATEMENT OF PROBLEM

Currently in USPS parcel sorting operations two people are necessary to face and code a parcel for sorting. It was anticipated that one of these positions could be eliminated or, alternatively, both positions could be devoted to coding, if an acceptable voice encoding system (VES) could be applied to the task. If acceptable throughput rates can be realized voice encoding offers several advantages over keyboard code entry. Paramount among these for the parcel sorting task is the ability for the operator to use his hands for other tasks while speaking. Thus the encoding operator could face parcels and code them for sorting simultaneously.

For voice encoding to be considered a feasible alternative to keyboard coding for parcel sorting or any other postal application two criteria must be met. First, the voice recognition hardware employed must meet the engineering requirements for the task. Second, it must be determined that the human voice is capable of the task demanded. A previous study employing the Numeric Speech Translator (NST) concluded that voice encoding can be performed without undue strain on the speaker provided that he is not asked to exceed a speaking rate of 30 codes (60 digits) a minute or 45 minutes of talking in any hour.[1] Determination of acceptance rates by the NST was incomplete and inconclusive due to equipment incompatibility. The results of this investigation suggest that voice encoding may be a feasible

1. NBS Report 10 588 The Human Voice As An Encoding Medium, June, 1970.

method of data input for postal sorting but do not offer conclusive evidence that acceptable throughput rates can be maintained.

The purpose of the work herein reported was to evaluate a VES in two phases. Phase I was designed to evaluate the VES using naive subjects in a simulated parcel sorting situation with ZIP sorting. Phase II was designed to test some of the limits of the man-machine system using highly motivated and trained subjects.

The VES employed in this study, chosen primarily on the basis of its availability, was developed to convert spoken commands into Binary Coded Decimal (BCD) output for interface with any digital input device. The operator simultaneously "teaches" the machine to recognize his voice and programs in the commands which he wishes to use by placing the machine in the "train" mode. Each command is spoken 5 times as the unit displays the number with which that command is to be associated. After all commands have been programmed the unit is switched to the "run" mode, in which it responds to recognized commands by displaying the appropriate number on the console display and by sending it in BCD form to the device to which it is interfaced. Specifications and photographs of the VES are presented in the Appendix.

2.0 PHASE I

2.1 Procedure

For all phases of this study stimuli consisted of ZIP coded letter mail presented via an LSM console. Subjects were required to read various sets of digits from the ZIP code into a Telex microphone interfaced with the VES and data recording equipment. A detailed account of the testing procedure

may be found in the study design plan in the Appendix. Deviations from this plan, as indicated below, were approved by the USPS Project Manager.

After a brief orientation and practice session, Phase I subjects encoded the last 2 digits of the ZIP Code under the various conditions of Presentation Rate and Work/Rest Cycle outlined in the diagram below:

Presentation Rate (Items/min)	Work/Rest Cycles (Min On/Min Off)		
	15/15	30/30	60/60
20	16	8	4
35	16	8	4

The numbers in the cells indicate the number of trials presented under each of the 6 conditions. The number of trials was equated for total time at the task. Subjects performed the voice encoding task approximately 6 hrs./day for 12 days.

A one minute sample of VES output and of operator response was obtained from every 5 minutes of operation. The measures obtained were: total system performance and subject reading errors. Total system performance was defined in terms of the percentage of mail pieces (items) which were correctly coded by the VES-operator system. To be correct both digits of the code must have been spoken correctly by the operator and recognized correctly by the VES.

2.2 Results

The mean percentage of items correctly coded under each of the 6 conditions is presented in Table I. Also shown are the individual VES recognition rates for each of the 4 subjects. The results of an analysis of variance performed on the data summarized in Table I indicated that overall a significantly greater percentage of items was correctly coded at a presentation rate of 20 items/min. (68.6%) than at 35 items/min. (48.4%). Also, total system

TABLE I

Total System Performance: Phase I
Mean Percent Correct Recognition of
Both Digits of 2 Digit Items

Subject	Presentation Rate Items/Min.	Work/Rest Cycle 15/15	30/30	60/60	Mean
1		54.9	48.6	60.5	54.7
2		66.9	69.0	74.1	70.0
3	20	70.4	72.3	81.9	74.9
4		67.8	76.1	80.8	74.9
Mean		65.0	66.5	74.3	68.6
1		27.3	22.7	27.0	25.7
2		54.2	42.6	60.4	52.4
3	35	66.9	56.9	74.2	66.0
4		37.8	39.2	70.9	49.3
Mean		46.6	40.4	58.1	48.4

performance was superior when operators were working under a 60/60 work/rest cycle than either a 15/15 or 30/30 cycle. This difference, however, is confounded with the order of presentation of the work/rest cycles. All subjects were tested on the shortest cycle first and the longest cycle last. That the 60/60 cycle was superior, therefore, can be an indication of learning and adapting to the task.

Total system errors, as reported above, do not differentiate between machine errors and operator errors. Thus an evaluation of operator reading errors was made. Table II shows the percentage of individual digits and items which were incorrectly coded as a result of reading errors. An item was scored as incorrectly coded due to reading error if either one or both of the digits was incorrectly read. Two types of reading errors were counted, omissions and misread digits. As is indicated in the Table, only a small portion ($<1\%$ overall) of the total system errors can be accounted for on the basis of operator reading errors. These errors generally decreased as a function of increased experience with the coding task.

2.3 Fatigue Test

In addition to coding stimuli under the six conditions resulting from the combination of 3 work/rest cycles and 2 presentation rates, Phase I subjects also participated in a test of fatigue factors in voice encoding. On 2 consecutive days each of the subjects was requested to continue voice encoding until they reported that they were extremely fatigued and could not continue. Subjects coded at a presentation rate of 20 LPM during one of these sessions and at 35 LPM during the other. As in other tests, a one minute sample of coding performance was taken from each 5 minute segment of testing.

TABLE II

Summary of Reading Errors: Phase I
Percent of Individual Digits and
Items Miscoded Due to Reading Errors*

Work/Rest Cycle	Items/Min.	Omissions		Misreads		Total	
		Individual Digits	Items	Individual Digits	Items	Individual Digits	Items
	20	.30	.54	.87	.91	1.17	1.45
15/15	35	.18	.32	1.18	1.24	1.36	1.56
	Mean	.24	.43	1.02	1.08	1.26	1.51
	20	.13	.21	.52	.56	.65	.77
30/30	35	.06	.13	.40	.46	.46	.59
	Mean	.10	.17	.46	.51	.56	.68
	20	.12	.23	.40	.45	.52	.68
60/60	35	.09	.18	.28	.30	.37	.48
	Mean	.10	.20	.34	.38	.44	.58
	20	.18	.32	.60	.64	.78	.97
Mean W/R	35	.11	.21	.62	.67	.73	.88
Mean All Conditions		.14	.26	.61	.66	.76	.92

*Note that for most conditions errors are expressed in fractions of 1%.

Table III summarizes the data obtained from the fatigue test. As is indicated in the table, average time at the task was greater at 20 LPM than at 35 LPM. Though this relationship did not hold for all subjects, coding time averaged 30 min. longer at the slower rate than at the faster rate. Total system performance, in terms of percent of items correctly coded was also superior at 20 LPM. Reading errors, presented in Table IV, again averaged $< 1\%$ and thus account for only a small portion of the total system errors.

During the fatigue tests, as in other phases of testing, the VES was retrained when, in the operator's and experimenter's judgment such retraining was warranted by a substantial decrement in the recognition of a specific command. The VES was retrained an average of 3 times during each fatigue test session, i.e. approximately every 37.5 min. Samples taken immediately after retraining showed an average 8.0% improvement in VES recognition over the sample taken immediately preceeding retraining. However, 41.7% of the samples taken after retraining indicated that retraining resulted in a decrement in performance. This was due to the necessity of retraining the entire vocabulary even if only one command was being rejected. The result was that other previously well recognized commands were often rejected.

Although it appeared that retraining was required due to the subjects' voice changing and not due to changes in the response of the VES, limited objective verification of this observation was attempted.

The methods used were limited to tape recordings which were found to be unsatisfactory. It is recommended that any future testing of VESs include the recording of the wave forms of the various words from each subject at the beginning of a test and at the time the system indicates a need for

TABLE III

Summary of Results from Fatigue Test: Phase I
Including: Time to Fatigue, Percent of Total
Items Correctly Coded, and Percent Reading Errors

LPM	Subject	Time to Fatigue	% Correctly Coded	% Reading Errors
	1	2 hrs. 5 min.	50.3	.79
	2	2 hrs. 30 min.	83.5	.64
20	3	3 hrs. 30 min.	82.6	.80
	4	3 hrs. 10 min.	76.9	1.17
	Mean	2 hrs. 49 min.	73.3	.85
	1	2 hrs. 5 min.	40.5	.67
	2	3 hrs.	67.9	.65
35	3	2 hrs. 30 min.	81.2	.40
	4	1 hr. 40 min.	59.7	2.27
	Mean	2 hrs. 19 min.	62.3	1.00

TABLE IV

Summary of Reading Errors: Phase I
Fatigue Test - Percent of Individual Digits
and Items Miscoded Due to Reading Errors

Items/Min.	Omissions		Misreads		Total	
	Individual Digits	Items	Individual Digits	Items	Individual Digits	Items
20	.25	.49	.31	.36	.56	.85
35	.26	.53	.41	.47	.67	1.00
Mean	.26	.51	.36	.42	.62	.92

"retraining". The optical recording oscillograph available in the USPS Laboratory is supplied with a galvanometer with a frequency response of 2500 Hertz. A unit of this type should be capable of presenting quantitative data for an analysis to evaluate the interface of the subjects and the hardware from psychological and engineering standpoints. Such data could indicate the best form for each subject to pronounce each of the vocabulary words and indicate the weak and strong areas of the VES being tested regarding its ability to recognize variations in wave patterns of input for each word.

3.0 PHASE II

3.1 Procedure

Phase II of the present study was initiated in order to better evaluate the limits and capabilities of the VES-operator system using more highly motivated and trained subjects than those employed in Phase I. NBS-HFL staff members who had served as experimenters in Phase I were used as subjects for Phase II. The first part of Phase II consisted of an evaluation of system performance with 3 subjects working under a 30/30 work/rest cycle and performing the following coding tasks:

last digit at 20 items/min.

last 2 digits at 20 items/min.

last 3 digits at 20 items/min.

last 4 digits at 20 items/min.

last 2 digits at 15 items/min.
with a reject and correct capability.

The purpose here was to determine the limits of code length entry at the rate indicated.

The second test undertaken in Phase II was a comparison of VES recognition rates under two levels of ambient noise.

3.2 Results

Table V shows the percent of items and individual digits correctly recognized by the VES under the coding conditions described above. These data were derived from 1 min. samples taken from each 3 min. of voice encoding. A comparison of Table V with Table I indicates that Phase II subjects realized considerably higher recognition rates than did Phase I subjects under comparable conditions. Working under a 30/30 work/rest cycle and coding the last 2 digits of the ZIP Code at 20 items/min., 81.7% of the items were correctly recognized by the VES during Phase II. VES recognition was only 66.5% for this condition in Phase I. This difference is likely attributable primarily to the higher motivation of Phase II subjects. In addition to the conditions for which data is presented in Table V, one subject attempted to speak all 5 digits of the ZIP Code at 20 items/min. This task proved nearly impossible for the subject to perform for more than only a few minutes. VES recognition was so low and reading errors so high that these data are not presented.

3.3 Cancel and Correct Capability

The most successful test of voice encoding performed in this study, in terms of the percentage of correct recognition, was that in which subjects spoke the last 2 digits of the ZIP Code at a rate of 15 LPM with a cancel and correct capability. In addition to training the VES to recognize the digits from 0-9, operators in this condition trained the machine to recognize

TABLE V

Summary of Phase II Subject Performance
as Measured in Terms of Percent Correct at 20 LPM

Percent Correct Recognition

Coding Task

	Last Digit	Last 2 Digits		Last 3 Digits		Last 4 Digits	
Subject		Items	Individual Digits	Items	Individual Digits	Items	Individual Digits
1	92.2	82.9	90.5	72.6	89.8	60.0	86.4
2	95.5	88.5	94.2	69.0	86.3	30.0	61.5
3	88.4	73.7	85.0				
Mean	92.0	81.7	89.9	70.8	88.1	45.0	74.0

the word "reject". When the operator observed a misrecognition on his feedback display he uttered the code "reject" and then re-entered the correct code. As is indicated in Table VI, this capability resulted in 97% correct recognition for the two subjects tested. If the reject-correct function had not been availble, these subjects would have realized only 86.6% correct recognition.

The results of voice encoding with a cancel capability provide substantial evidence in this study that voice encoding may be a feasible alternative to keyboard entry. Higher recognition rates may be obtained if another method of signaling "cancel" were available. This should result in both increased recognition rates and the capability to increase the speed of presentation.

3.4 Error Analysis

The data obtained from subjects encoding a single digit at 20 items/min. were submitted to an error analysis to determine if recognition errors were systematic, i.e., if specific digits were consistently misrecognized, or if errors were distributed randomly among the ten digits. Table VII indicates the frequency that any given digit was recognized as any other digit. "Reject" refers to the frequency with which a spoken digit resulted in no output from the VES. This composite of errors, based on the performance of 3 subjects, indicates that the most consistent errors occurred with the digits 1, 5, 6 and 8. The cells in Table VII which are circled indicate the misrecognition errors which were most common.

Operator reading errors for Phase II subjects were calculated from samples taken on magnetic tape. Table VIII shows these errors, both omissions and misreads, for both individual digits and items. Under most conditions

TABLE VI

Mean Percent Correct When Phase II Subjects
Were Able to Cancel and Correct Responses

Percent Correct Recognition

Coding Task

15 LPM

Subject	Last 2 Digits (with correction capability)	
	A[1]	B[2]
1	85.4	96.6
2	87.7	97.5
3		
Mean	86.6	97.0

1-Percentage without corrections
2-Percentage after corrections

Digit Spoken	# Correct	VES Response Frequency											total errors	total spoken	% errors
		Reject*	0	1	2	3	4	5	6	7	8	9			
0	428	4	X	1	1	5	-	-	2	1	-	-	14	442	3
1	350	(29)	1	X	-	-	1	(14)	-	3	1	6	55	405	14
2	345	-	8	-	X	3	-	1	2	1	2	-	17	362	5
3	301	5	9	4	5	X	-	-	5	2	3	2	35	336	10
4	313	7	1	3	1	-	X	2	-	1	-	-	15	328	4
5	281	5	2	(17)	4	1	1	X	-	-	-	7	37	318	12
6	358	5	5	-	(14)	7	1	1	X	7	(33)	-	73	431	17
7	314	3	4	4	12	4	-	1	6	X	3	3	38	353	11
8	279	4	-	-	(13)	5	-	-	10	3	X	3	38	317	12
9	322	4	-	2	1	1	-	4	-	4	-	X	16	338	5
Total	3291	66	30	31	51	26	3	23	25	22	42	21			

*"Reject" refers here to the frequency with which a spoken digit resulted in no output from the VES.

TABLE VII

Specific Digit Error Analysis for 3 Subjects:
Phase II - 20 items/min., last digit only

TABLE VIII

Mean Percent Reading Error: Phase II

Items/ Min.	Digits Encoded	Omissions		Misreads		Total	
		Individual Digits	Items	Individual Digits	Items	Individual Digits	Items
20	Last Digit	0	0	0	0	0	0
	Last 2	.57	1.14	.17	.20	.74	1.34
	Last 3	.46	.98	.92	.92	1.38	1.90
	Last 4	2.40	3.66	4.24	4.24	6.64	7.90
	Last 2 with Correction	.05	.10	1.41	1.41	1.46	1.51

reading errors were made on less than 2% of the items spoken. As the number of digits per item increased, however, the percentage of reading errors also increased. A substantial number of errors was made when subjects spoke 4 digits of the ZIP Code.

3.5 Noise Test

The second test undertaken under Phase II was a comparison of VES recognition under two levels of ambient noise. The stimuli for all tests reported thus far were presented via an LSM console. Since the noise generated by the LSM would not necessarily be present in a field application of voice encoding a comparison was made of VES recognition with the LSM on and with the LSM off. Six subjects were tested under both "noise" and "no noise" conditions. The task consisted of reading 100 single digits from a printed list. Each subject performed the task 3 times under both conditions. Table IX shows the results of this test. As is clearly indicated in the table no significant difference was observed between VES recognition rates under the two levels of ambient noise. The results of a t-test performed on these data verify this lack of difference ($t = .19$, $df = 5$).

4.0 CONCLUSIONS AND RECOMMENDATIONS

The present study was performed to evaluate the feasibility of employing voice encoding in postal sorting applications. Although the data obtained from this study apply directly only to the specific unit employed, many observations were made relevant to voice encoding in general.

The feasibility of voice encoding as an alternative to keyboard entry was most clearly demonstrated during Phase II when the subjects had a cancel and correct capability available. The subjects obtained a 97% correct level

TABLE IX

VES Recognition Under Two Conditions of Ambient Noise - Mean Correct for 3 Trials of 100 Digits Each, Self-Paced Mean Percent Recognition

Subject	Noise	No Noise
1	90.3	87.0
2	86.3	92.0
3	88.0	86.7
4	94.7	88.7
5	94.9	91.0
6	81.7	88.0
Mean	89.3	88.9

of machine recognition when coding 2 digits at 15 LPM. This task should be equivalent to some sorting operations. It is anticipated that even higher recognition and throughput rates could be obtained if an alternative method could be used to indicate cancel.

The best VES recognition rates obtained from Phase I (naive) subjects were observed under conditions of voice encoding at a presentation rate of 20 items/min. and a 60/60 work/rest cycle.[1] The mean recognition rate under these conditions was, however, only 74.3%. Clearly these data do not support a recommendation to employ voice encoding with the present system. Much was learned during Phase I, however, which aided in obtaining the much improved performance evidenced in Phase II.

The most important data obtained from Phase I testing relates to the large differences in VES recognition rates as a function of operator voice characteristics. Recognition rates for Phase I subjects ranged from 22.7% to 81.9%. The subject who consistently produced the lowest rates was also the most variable in his speech patterns. This subject had an accent which changed markedly with the onset of fatigue. Conversely, the subject who consistently produced the highest recognition rates was, in the experimenter's estimate, the most consistent in terms of voice characteristics. Consistency here was a subjective evaluation of the experimenter.

Another important determinant of VES recognition rates appears to be operator motivation. Phase II subjects, NBS-HFLC staff members, clearly had a greater interest in the study than Phase I subjects and, hence, performed at a level superior to the naive subjects for the same condition (Mean percent correct - Phase I = 66.5%, Phase II = 81.7%). This higher level of motivation likely functioned to offset the tedious nature of the voice encoding task

1. That this work/rest cycle was superior is likely due to learning.

employed in this study. If voice encoding is to be considered for postal application, methods must be developed to provide enough incentive to overcome the boredom inherent in the task. Since the field use of VES would be job related, a motivation level higher than that of experimental subjects could be assumed, but additional incentives would be advantageous.

It should be remembered that for all phases of testing in the present study, voice encoding was performed under conditions of machine pacing. Operators had no control over the rate at which items were presented for coding. It is conceivable, but at present untested, that equal or superior production rates could be achieved under conditions of self pacing. In order to implement self pacing, however, a unique command or some other method of indicating end of message would have to be incorporated in the task. Given this requirement, self pacing may still be feasible since it would eliminate the problem, often observed in the present study, of operators "getting behind". When this occurred the operators attempted to "catch up". This generally resulted in a series of misrecognized commands.

4.1 Recommendations: Equipment Design

Observations made during the course of this study suggest several features a voice encoding system should have for successful postal application. The following are recommended:

1. Any system considered by USPS should have the capability of training individual commands without retraining the entire set. As a test period progressed in the present study, specific commands often were not recognized. When this was observed, the unit was retrained. The training procedure for the unit tested, however, requires that the entire vocabulary of commands be

retrained. This often resulted in deterioration in the recognition rates of other previously highly recognized commands.

2. A "reject" or "cancel" capability accessed by other than a unique verbal command is recommended. As noted in Phase II results, the best recognition rates were observed when a "cancel and correct" function was included. Higher production rates may have been reached with this capability if some other mode of signaling this function were available.

4.2 Recommendations: Task Parameters

1. Length of Code

The results of the present study indicate that the maximum code length feasible for voice encoding (employing the unit tested) should be 2 digits/item at a presentation rate of 20 items/min. Higher production rates may be obtained with an efficient cancel and correction capability.

2. Work/Rest Cycle

Of the work/rest cycles tested in the present study, performance was best under the 60/60 cycle. This was likely due, however, to learning effects. Observations made during the course of the study and in previous investigations indicate that the optimum work/rest cycle for a full time voice encoding task would be 30/10 or 45/15.

3. Motivation

Because of the tedious nature of machine paced voice encoding, methods of providing additional incentives should be explored.

4. Effects of Ambient Noise

No differences in VES recognition rates were observed in the present study as a function of ambient noise level. Comparisons were made, however, under only two levels of noise, i.e. LSM console on, and LSM off. VES recognition may be affected to a greater degree in a live sorting situation. This possibility should be investigated.

5. Field Test

It is recommended that a field test, or a laboratory test employing a PSM simulator, be performed in which VES operators code "live" parcels. The handling of parcels and freedom of movement provided in such a test may significantly reduce the tedium associated with the voice encoding task and thus improve performance.

6. Two-man Operation for Parcel Sorting

USPS should consider the feasibility of both the facer and keyer in present parcel sorting operations adopting voice encoding. If both of these personnel encode parcels a rate of 40 parcels/min. could be maintained.

7. Intermittent Use of Voice Encoding

USPS should consider employing voice encoding of parcels during slack periods. If a VES were available during periods when few parcels are available for sorting, a single operator could perform both the facing and coding operations. At peak periods sorting could revert back to current keyboard sorting procedures.

Although the results of the present study do not present overwhelming evidence in favor of immediate adoption of voice encoding by USPS, there is evidence that voice encoding could be a feasible alternative to keyboard coding for some sorting applications. Further investigations of voice encoding systems other than the unit tested in the present study is suggested to adequately evaluate this possibility. In addition, more detailed engineering evaluation concerning the optimum recognition system should be initiated and surveys of USPS tasks should be carried out to determine the areas where voice encoding could best be applied.

APPENDIX A

Study Design Plan

A DESIGN PLAN FOR AN EVALUATION OF VOICE ENCODING SYSTEMS

1 March 1972

Prepared by:

R. D. Petersen
J. J. Persensky

A Design Plan for an Evaluation of Voice Encoding Systems

A. BACKGROUND

Problem. The Human Factors Laboratory has been requested to evaluate various voice encoding systems (VES) in order to determine the feasibility of utilizing voice encoding in a parcel sorting application. It is anticipated that a VES could reduce the cost of parcel sorting by both reducing the error rates inherent in keyboard encoding and by possibly enabling a single operator to perform both the facing and coding tasks. However, whether any particular voice encoding system can realize these potential savings will depend to a great extent upon that system's limitations in its interface with the human operator. It is necessary that the HFL have the use of VES units for at least three months in order to subject them to the program of evaluation described in the present document.

The Equipment. One VES, which has been developed by , will be the first system to be evaluated at the NBS-HFL. This particular VES is a voice encoding system developed to convert spoken commands into BCD (Binary Coded Decimal) output for interface with any digital input device. Total vocabulary size varies from one model to another up to a maximum of 40 commands. The operator simultaneously "teaches" the machine to recognize his or her voice and programs in the commands which he wishes to use by placing the machine in the "train" mode and speaking each command as the machine displays the number with which he wishes that command to be associated. After being trained, the machine is

switched to the "run" mode, in which it responds to a recognized command by displaying the appropriate number on the console display and feeding it in BCD form to the device to which it is interfaced. According to the manufacturer's specifications, each command must be one second or less in duration and commands must be separated by at least 250 ms.

It is anticipated that other VES to be evaluated will have similar characteristics so that the study design herein presented will suffice for evaluation of other systems.

Application. Most USPS parcel sorting locations utilize a conveyor system which carries the packages to as many drop off points as there are primary sorts. The system is controlled by an encoding operator who keys the scheme destination or ZIP Code of each parcel into a PSM keyboard as the mail piece passes his position. Currently each sorting station is manned by two operators, one of whom, at any given time, encodes while the other acts as a "facer", turning each parcel so that the keyboard operator can read the address. It is common for the two operators to trade jobs approximately every fifteen minutes. It is believed that the use of a VES might eliminate the coding job at each station since the facer could perform both tasks. It is also anticipated that voice encoding will result in fewer errors than keyboard encoding. These two factors should result in considerable savings to the USPS in terms of reduction in both personnel and errors, thus also reducing cost.

A number of factors related to VES may affect the balance of cost, however, and insofar as they can be simulated in a laboratory setting, these factors will be taken into account in the proposed evaluation. The factors are: error rate (total system, operator, and machine), training, pacing of work, work/rest cycle (fatigue), feedback, and worker attitudes.

B. METHODOLOGY

A two phase experiment is proposed. Phase I will employ trained VES users selected from Dimension employees. The purpose here is: to familiarize the HFL and USPS staff with the VES, to develop a criterion error rate under ideal conditions and to develop optimized methods of training operators. Postal Service personnel and NBS supplied personnel (both male and female) will serve as subjects in Phase II. These experiments will be designed to evaluate the VES with a simulation of a ZIP Code sort operation by means of the levels of the various independent variables.

1. Independent and Dependent Variables: The independent variables will include: stimulus presentation rate, work/rest cycle, and visual feedback. Three stimulus presentation rates will be examined: 35, 40, 45 items/min. Work/rest cycles of 15/15 min., 30/30 min., and 60/60 min. will be tried in a partial evaluation of fatigue. Two levels of visual feedback will be used: Feedback Present (P) and Feedback Absent (NF).

The dependent variables will include production and error rates, measures of fatigue and training time. Three types of error counts will be made: total system, machine error, and operator error. Numbers of rejects and cancels will also be determined. No truly objective measure of the effects of fatigue can be employed but approximations will be derived based on performance differences between the various work/rest cycles and on the subjective evaluations of fatigue as given by the operators. Training time will be measured by time to reach the criterion of performance established by trained VES operators and performance changes after trials for the other operators.

2. Design: An incomplete within-subjects 3 (machine pace) x 3 (work/rest) x 2 (feedback) design will be employed, with three levels of machine pacing (35, 40 and 45 items per minute); three work/rest cycles (15 on/15 off, 30 on/30 off, 60 on/60 off); and two levels of feedback (present and absent). The order in which each subject will receive the various combinations of treatments will be counterbalanced. The design is diagrammed below:

Items per Minute	Work/Rest Cycle (Min. On/Min. Off)			
	Feedback			No Feedback
	15/15	30/30	60/60	15/15
35	16	8	4	16
40	16	8	4	16
45	16	8	4	16

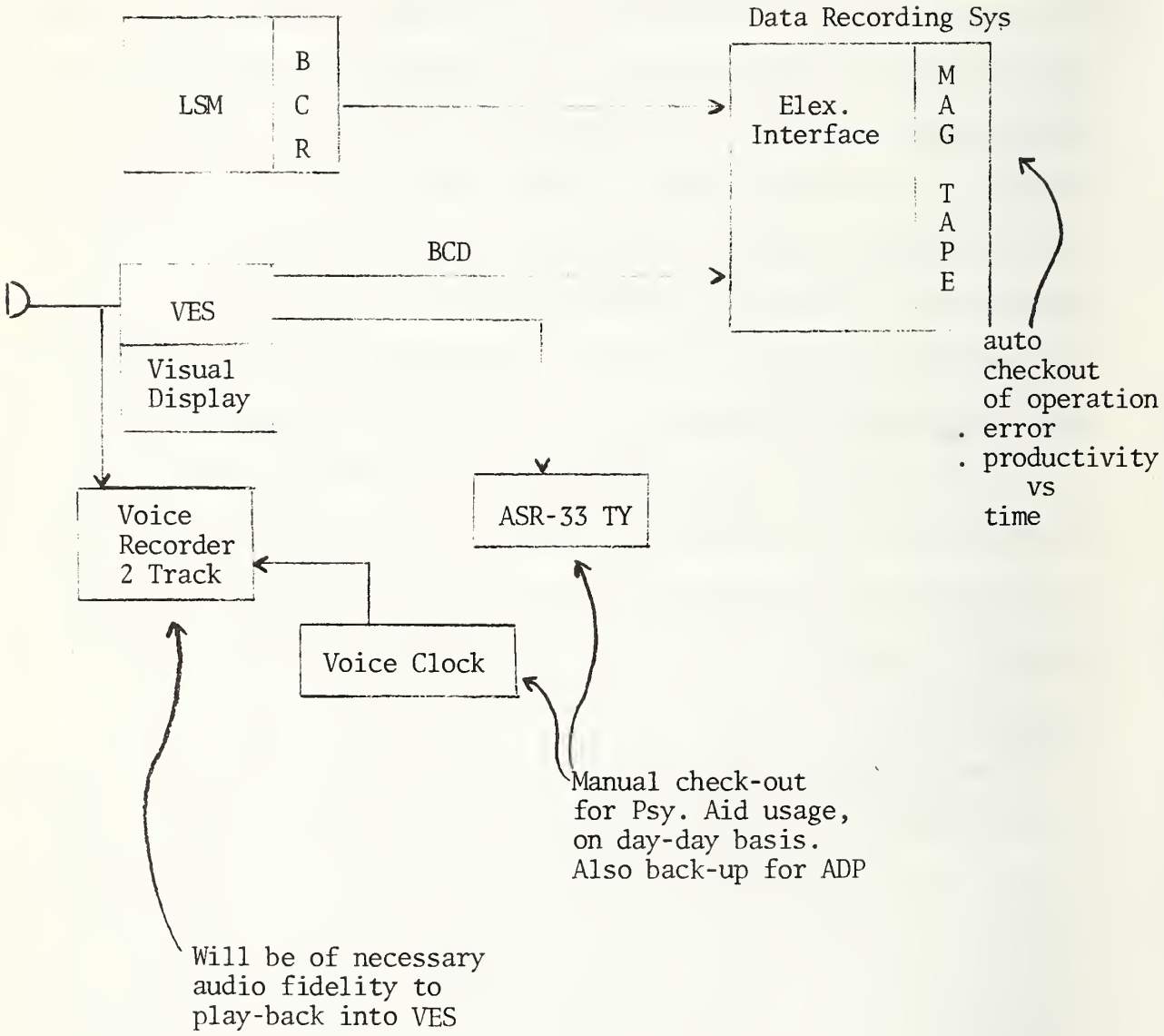
Numbers in the cells (16, 8, 4) indicate number of trials in each of these conditions. Trials will vary in order to hold test time and number of addresses presented constant across work/rest cycles.

The design is incomplete in that only one of the work/rest cycles (15/15) will be examined in the No Feedback condition. Thus the comparison for this variable can only be made with the analogous groups in the Feedback condition.

3. Subjects: will be requested to supply at least 1 trained VES operator for approximately 4 hrs./day for the one week duration of Phase I. For Phase III the USPS is requested to supply 4 male P.S. personnel for 3 weeks each at 8 hrs./day, and NBS will supply 4 female subjects for 3 weeks each at 8 hrs./day.

4. Apparatus: The apparatus will consist of the VES with accessories including an appropriate microphone, a digital printer to be interfaced to the output of the VES, a magnetic tape recorder connected to receive the same voice signal which is fed into the VES, and an LSM trainer. The equipment configuration is shown in Figure 1. The microphone is inputted to the VES and to a voice recorder. The VES output, a digital code in ASC II, will be inputted to a hard copy printer for data reduction. The voice tape will be used as a check of input for later data analysis. Subjects will read addressed mail displayed at various speeds by the LSM trainer. The LSM data recording system will record bar code data on tape and this information will constitute a reproduction of the actual stimuli presented to the operator. The list of ZIP Codes so developed will later be compared with operator and VES output. Other necessary materials will include letter mail stimuli, rating scales to be administered periodically to measure fatigue and other subjective reactions, and attitude questionnaires to be administered twice to each subject.

5. Procedure: Subjects in both phases of the study will be treated in as nearly the same manner as possible. After instructions as to the purpose of the experiment and how to use the VES, the subjects will be asked to "train" the VES in the required vocabulary. Training techniques will be acquired by the experimenters during Phase I. Subjects will then be requested to read Zip coded letter mail and to speak aloud 2 digits from the ZIP Code into a microphone. Each of the subjects will be tested under the 12 conditions determined by the combinations of the 3 independent variables presented in a balanced order. That is, they will participate in the experiment under the several levels of the conditions of pacing, work/rest cycle, and feedback.



VOICE ENCODING TEST CONFIGURATION

Figure 1

More specifically:

Phase I: The VES contractor will supply 1 subject for 4 hours per day for 5 days. This subject will participate in each of the 9 conditions determined by the interaction of the variables of work/rest cycle and mail piece presentation rate. He will carry out:

1) 4 trials at each presentation rate for the 15/15 min. work/rest cycle;

2) 2 trials at each rate for the 30/30 min. cycle;

3) and 1 trial for each rate within the 60/60 min. cycle.

Data from this work will be used to establish the production and error criteria as well as methods for training the VES.

Phase II: The 8 USPS and NBS subjects will be tested in pairs for 8 hours per day for 14 days. The first 12 days will be reserved for testing within the 12 conditions of the combinations of the 3 independent variables. The subjects will view the mail piece and speak into the microphone the last 2 digits of the ZIP Code for the amount of time required by that experimental condition. The last two days of the experiment are intended as an intensive test of fatigue factors. On these days the subjects will be asked to continue voice coding until they report that they are extremely fatigued. Two presentation rates will be tested, 35 and 45 LPM.

Presentation rate and feed back variables will be randomized within and between subjects and days. Work/rest cycles will be randomized between and within subjects and between days but not within days, i.e., the work/rest cycle must be kept constant during any particular day.

The total duration of Phase II will be 12 weeks.

C. DATA COLLECTION

For analysis of the measures of production and accuracy data will be collected from 4 sources: VES output on 7 track, Sigma II compatible tape; VES direct printout, voice tape, and the stimulus list. One minute samples of these VES direct-printout data will be collected from within each 5 min. block of test time. 100% data capture will be obtained from the magnetic tape recordings. Utilizing this sampling procedure from the VES direct printouts will result in approximately 13,440 data points for the 35 LPM presentation rate, 15,360 data points for the 40 LPM presentation rate, and 17,280 data points at the 45 LPM rate.

Other data collected will be subjective responses to rating scales and attitude questionnaires.

D. DATA REDUCTION AND ANALYSES

Data of most importance to the present study are the error rates. Three types of errors will be derived from comparisons between the stimulus list, the voice recording tape, and the digital printout. Deviations of the VES output from the stimulus list represent total system errors. Deviations of the voice tape from the stimulus list represent operator errors. Deviations of the printout from the voice tape represent machine errors. Error rates, i.e., relative frequencies, will be used for comparisons among conditions because the number of stimuli in the various conditions will not be the same. Other data which will be collected include: number of cancels, number of rejects, number of times required to "retrain" VES after initial settings and time to reach the established criterion. A subjective rating scale will also be administered to the subjects after each condition to determine personal feelings of fatigue or discomfort.

Finally, a questionnaire will be constructed to assess each subject's attitudes toward VES and its possible adoption by USPS. The questionnaire will be administered to each subject at the start of the study and after he has finished his participation in the experiment. The questionnaire will be designed to measure how positively or negatively postal clerks feel about VES, whether they think it should be adopted, and how the task structure of the parcel sorting location should be changed to accommodate it. Appropriate measures of central tendency and dispersion will be calculated for the various measures and presented in tabular and graphic formats along with any statistical analyses deemed necessary.

F. FUTURE EVALUATIONS

It is assumed that the study design herein contained will serve as a basic medium around which evaluations of other VES can be planned.

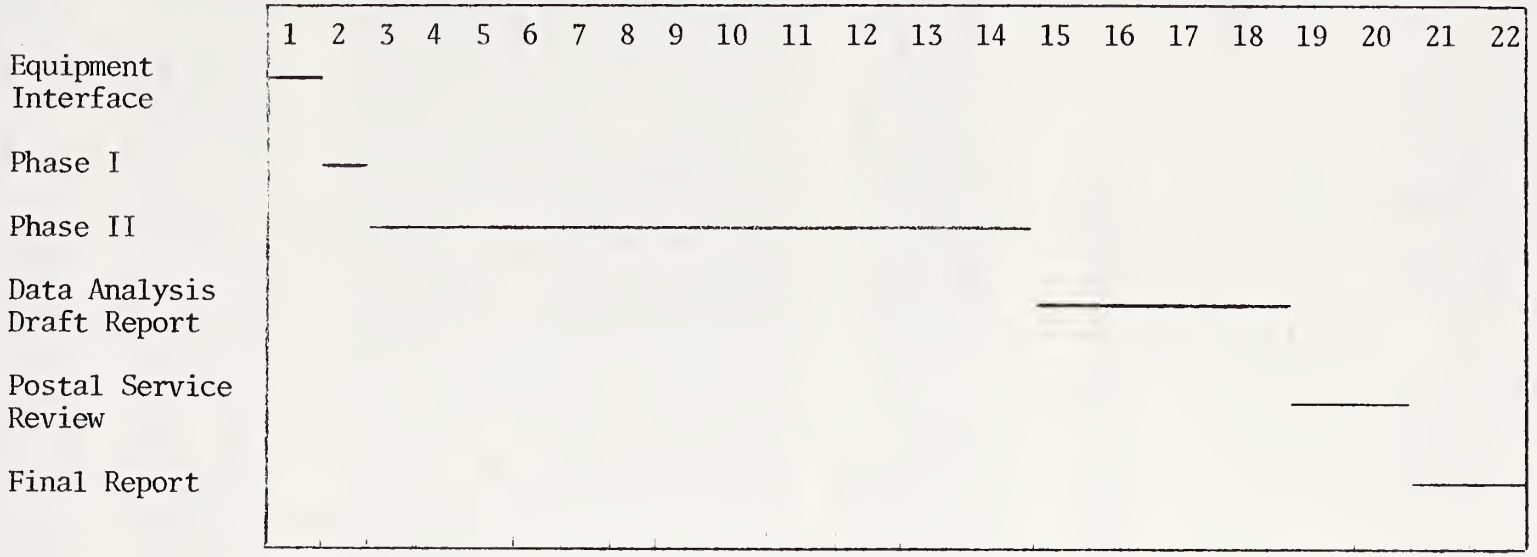
TABLE I

A block diagram is attached* showing the test set-up to be used. The following equipments are involved:

1) Voice Encoding System
2) LSM for handling letter stimulus material and reading bar code
3) Voice recorder
4) ASR-33 TY
5) #1 Data Recording System with Mag Tape
6) Voice clock.

* See Figure 1.

Projected Schedule for
An Evaluation of Voice Encoding Systems



U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

Date: August 29, 1972

To: J. K. Giles

From: J. Persensky

Subject: Continuation of Voice Encoding System Evaluation

As mentioned in the attached memo (16 August 1972), further testing of the according to the original study design (3 February 1972) has been curtailed at USPS request. In order to best evaluate the limits of operation, the following procedures were suggested and have been verbally approved by Mr. Giles.

Highly motivated subjects (NBS-HFLC staff) will be tested under a series of conditions in order to establish performance curves. The subjects will begin testing at 20 LPM, reading only 1 digit of the ZIP Code. After the criterion of 2 trials above 95% accuracy or 8 non-criterion trials, the subjects will be tested for performance reading 2 digits to criterion followed by reading 3, 4, and 5 digits. A trial will consist of a 30 minute work period; thus each condition will be tested for a maximum of 240 minutes. Error data will be manually collected for 1-minute samples out of every 3 minutes, so that 10 one-minute samples will be collected for every 30 minute trial. Time data and more extensive error data will be collected on a 1-out-of-5 minute schedule via the automatic data recording system. The manually collected data will be reported daily, while the automatically recorded data will be submitted when available. Presentation rate will also be increased or decreased depending on the outcome of the earlier trials.

A second task will be to test the effects of LSM noise on machine recognition. This will be accomplished by testing a number of subjects under noise and no noise conditions. The task will be to read 100 single digits from a printed list. Each subject will perform the task 3 times in each condition. Differences in error rate will be determined.

In order to simulate a field application of the , an attempt will be made to reach a 100% accuracy criterion when a cancel capability is added. The task will be reading 2 digits at various speeds between 10 and 20 LPM. This task should be similar to that of current sack sorting operations.

The final task will be a machine evaluation of the repeatability of the . In order to accomplish this a number of voice tapes will be developed (with known accuracy) which can be played back to the electronically. Two conditions will be tested: (1) with original recognition training only, and (2) with recognition training on every playback.

J. Persensky
Director, NBS-HFLC

Attachment

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

Date: August 16, 1972

To: J. Persensky

From: Val Pezoldt

Subject: Evaluation of VES

Preliminary evaluation of data obtained from the Evaluation of Voice Encoding Systems give strong indications that the system is not performing at a level appropriate for utilization by USPS. Recognition rates for the first two subjects average approximately 80% and 70% for stimulus presentation rates of 20 LPM and 35 LPM respectively.

In an effort to overcome this problem, it has been agreed (J. Giles, D. Cornog, J. Persensky, V. Pezoldt) to terminate the study as designed, on 18 August 1972, following the completion of 4 subjects' participation. As an alternative to following the design plan, personnel from NBS-HFLC will receive extensive experience as VES operators to determine if acceptable recognition rates (i.e., 95 - 100%) can be reached and maintained. Initially, a presentation rate of 20 LPM will be employed. If acceptable performance can be maintained at this speed, faster rates may be used.*

If acceptable recognition rates can be maintained by a single, experienced operator, effort will be directed towards determining training procedures to use with naive subjects.

*No attempt will be made during this testing to assess the effect of various work/rest cycles.

V. Pezoldt

APPENDIX B

Voice Encoding System Specifications

DIRECT VOICE CONTROL OF MACHINES

This voice encoding system converts an acoustic input to a digital code which can be used, with proper interfacing, to enter data into a computer, retrieve stored information, or control a machine operation. The voice encoding system can be trained to recognize the human voice for a wide range of speakers, vocabularies and acoustic environments.

SPECIFICATIONS

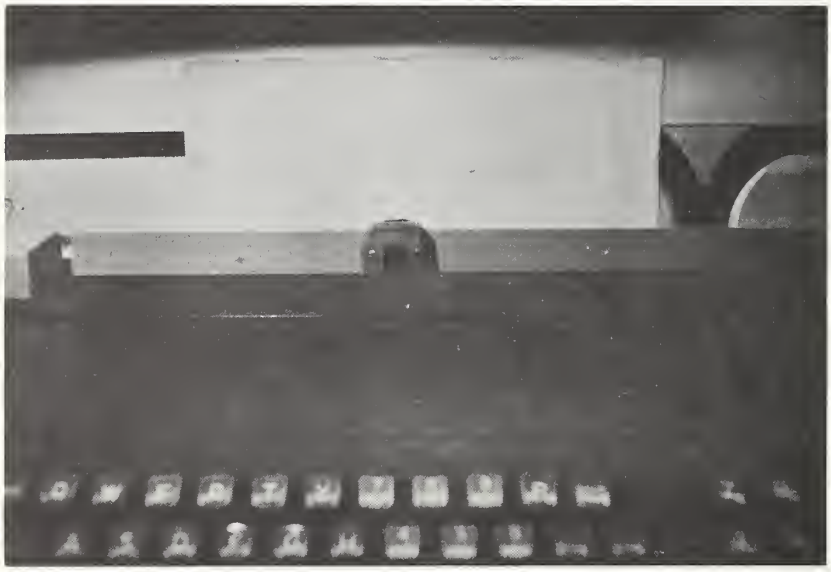
Input	24 spoken commands each approximately one second long
Output	Binary-coded-decimal (BCD) index of command Visual display
Operation	Trainable on line Training time less than 10 seconds per command Response time approximately 20 milliseconds Command spacing 250 milliseconds minimum
Physical	Power 115 v ac, 60 Hz, 200 w Size 8-3/4 x 19 x 23-1/2 inches Weight 85 pounds
Options	Telephone interface Voice response Additional vocabulary Foot switch Special output interface

APPENDIX C

Photographs

Voice Encoding System and
data recording equipment

Subject at LSM console voice encoding
ZIP Code digits



LED feedback display of
VES output

FORM NBS-114A (1-71)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 73-131	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

Human Factors Evaluation of a Voice Encoding System

5. Publication Date

March 7, 1973

6. Performing Organization Code

7. AUTHOR(S)

V. J. Pezoldt, J. J. Persensky, NBS-HFL Staff

8. Performing Organization

NBSIR 73-131

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.

4314467

11. Contract/Grant No.

715150S

12. Sponsoring Organization Name and Address

Postal Service Laboratory
Research Department
U. S. Postal Service

11711 Parklawn Drive
Rockville, Md. 20852

13. Type of Report & Period Covered

Final

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

The feasibility of employing voice encoding as an alternative to keyboard entry for postal sorting application was investigated in two phases. Phase I was a simulated parcel sorting situation. Phase II was performed to determine man-machine interface limitations. The results of this study led to the following conclusion and recommendations. Voice encoding would be of limited value to high speed parcel sorting unless it was used only during slack periods or the current facer and keyer both use a VES. Voice encoding could be applied to other sorting tasks. USPS tasks should be analyzed for the possiblilty of such application. Other voice encoding systems should be evaluated. These systems should have single command training capability, a cancel capability accessed by other than a unique verbal command, and should require infrequent system training.

17. KEY WORDS (Alphabetical order, separated by semicolons)

Feasibility; keyboard alternative; man-machine interface; postal sorting; voice encoding

18. AVAILABILITY STATEMENT

[X] UNLIMITED.

[ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NTIS.

19. SECURITY CLASS (THIS REPORT)

UNCLASSIFIED

21. NO. OF PAGES

55

20. SECURITY CLASS (THIS PAGE)

UNCLASSIFIED

22. Price